# EpiBuffer: Spatially-Varying Buffer Radii Model for Estimating Spatial Heterogeneity in Exposure Buffers and Risk

## SpatialBuffers Statistical Model

$$Y_i\left(\mathbf{s}_j\right)|\mu_i\left(\mathbf{s}_j\right), \boldsymbol{\zeta} \overset{\text{ind}}{\sim} f\left(y \mid \mu_i\left(\mathbf{s}_j\right), \boldsymbol{\zeta}\right),\ j=1,...,m,\ i=1,...,n_j;$$
$$g\{\mu_i(\mathbf{s}_j)\} = \mathrm{O}_i(\mathbf{s}_j) + \mathbf{x}_i(\mathbf{s}_j)^{\mathrm{T}}\boldsymbol{\beta} + \mathrm{z}\left(\mathbf{s}_j; \delta\left(\mathbf{s}_j\right)\right)\theta\left\{\delta\left(\mathbf{s}_j\right)\right\}.$$

## Likelihood Options

- Binomial likelihood with logit link function (likelihood_indicator = 0):

$$Y_i(\mathbf{s}_j)|p_i(\mathbf{s}_j) \overset{\text{ind}}{\sim} \text{Binomial}\left(\tilde{n}_i(\mathbf{s}_j), p_i(\mathbf{s}_j)\right), \text{logit}\left\{p_i(\mathbf{s}_j)\right\} = \mathbf{x}_i(\mathbf{s}_j)^{\mathrm{T}}\boldsymbol{\beta} + \mathrm{z}\left(\mathbf{s}_j; \delta\left(\mathbf{s}_j\right)\right)\theta\left\{\delta\left(\mathbf{s}_j\right)\right\};$$

- Gaussian likelihood with identity link function (likelihood_indicator = 1):

$$Y_i(\mathbf{s}_j) = \mathbf{x}_i(\mathbf{s}_j)^{\mathrm{T}}\boldsymbol{\beta} + \mathrm{z}\left(\mathbf{s}_j; \delta\left(\mathbf{s}_j\right)\right)\theta\left\{\delta\left(\mathbf{s}_j\right)\right\} + \epsilon_i,\ \epsilon_i|\sigma^2_\epsilon \overset{\text{iid}}{\sim} \mathrm{N}\left(0, \sigma^2_\epsilon\right);$$

- Negative binomial likelihood with logit link function (likelihood_indicator = 2):

$$Y_i(\mathbf{s}_j)|r, p_i(\mathbf{s}_j) \overset{\text{ind}}{\sim} \text{Negative Binomial}\left(r, p_i(\mathbf{s}_j)\right), \text{logit}\left\{p_i(\mathbf{s}_j)\right\} = \mathrm{O}_i(\mathbf{s}_j) + \mathbf{x}_i(\mathbf{s}_j)^{\mathrm{T}}\boldsymbol{\beta} + \mathrm{z}\left(\mathbf{s}_j; \delta\left(\mathbf{s}_j\right)\right)\theta\left\{\delta\left(\mathbf{s}_j\right)\right\}.$$

## Exposure Definitions

- Counts (exposure_definition_indicator = 0):

$$\mathrm{z}\left(\mathbf{s}_j; \delta\left(\mathbf{s}_j\right)\right) = \sum_{k=1}^{h} 1(\mathrm{d}_{jk} \le \delta\left(\mathbf{s}_j\right));$$

- Spherical (exposure_definition_indicator = 1):

$$\mathrm{z}\left(\mathbf{s}_j; \delta\left(\mathbf{s}_j\right)\right) = \sum_{k=1}^{h} 1(\mathrm{d}_{jk} \le \delta\left(\mathbf{s}_j\right))\left(1 - 1.5\left\{\frac{\mathrm{d}_{jk}}{\delta\left(\mathbf{s}_j\right)}\right\} + 0.5\left\{\frac{\mathrm{d}_{jk}}{\delta\left(\mathbf{s}_j\right)}\right\}^3\right);$$

- Presence/absence (exposure_definition_indicator = 2):

$$\mathrm{z}\left(\mathbf{s}_j; \delta\left(\mathbf{s}_j\right)\right) = \min\left(\sum_{k=1}^{h} 1\left(d_{jk} \le \delta\left(\mathbf{s}_j\right)\right), 1\right);$$

  - $\mathrm{d}_{jk}$: Distance between location $\mathbf{s}_j$ and point source $\mathbf{c}_k$;
  - $h$: Total number of point sources.

## Spatially-varying Radii and Exposure Effect Parameters

- Radii:

$$\Phi^{-1}\left(\frac{\delta\left(\mathbf{s}_j\right) - a}{b - a}\right) = \mathbf{w}(\mathbf{s}_j)^{\mathrm{T}}\boldsymbol{\gamma} + \phi(\mathbf{s}_j);$$

  - $\boldsymbol{\phi}^{\mathrm{T}} = \left(\phi\left(\mathbf{s}_1\right), \ldots, \phi\left(\mathbf{s}_m\right)\right)|\rho_\phi \sim \mathrm{MVN}\left(\mathbf{0}_m, \Sigma\left(\rho_\phi\right)\right)$, $\Sigma_{ij}\left(\rho_\phi\right) = \exp\left\{-\rho_\phi||\mathbf{s}_i - \mathbf{s}_j||\right\}$;
  - Gaussian predictive process option also available for large $m$;

- Effects:

$$\theta\left\{\delta\left(\mathbf{s}_j\right)\right\} = \sum_{l=0}^{p}\left\{\frac{\delta_j\left(\mathbf{s}_j\right) - a}{b - a}\right\}^l \eta_l.$$

## Prior Information

$\beta_j \overset{\text{iid}}{\sim} \text{N}\left(\mathbf{0}, \sigma^2_\beta\right),\ j = 1, \dots, p_x;$

- $p_x$: Length of $\mathbf{x}_i$ vector (same for all $i$), including an intercept;
- Default setting: $\sigma^2_\beta = 100^2$;

$\gamma_j \overset{\text{iid}}{\sim} \text{N}\left(\mathbf{0}, 1\right),\ j = 1, \dots, p_w;$

- $p_w$: Length of $\mathbf{w}\left(\mathbf{s}_j\right)$ vector (same for all $j$), including an intercept;

$\eta_j \sim \text{N}\left(0, \sigma^2_\eta\right),\ j = 1, \dots, p;$

- Default setting: $\sigma^2_\eta = 100^2$;

$\rho_\phi \sim \text{Gamma}\left(a_{\rho_\phi}, b_{\rho_\phi}\right);$

- Default setting: $a_{\rho_\phi} = b_{\rho_\phi} = 1$;

Gaussian specific: $\sigma^2_\epsilon \sim \text{Inverse Gamma}\left(a_{\sigma^2_\epsilon}, b_{\sigma^2_\epsilon}\right);$

- Default setting: $a_{\sigma^2_\epsilon} = b_{\sigma^2_\epsilon} = 0.01$;

Negative binomial specific: $r \sim \text{Discrete Uniform}\left[a_r, b_r\right];$

- Default setting: $a_r = 1, b_r = 100$.

## Default Initial Values

- $\beta_j = 0$ for all $j$;
- $\gamma_j = 0$ for all $j$;
- $\eta_j = 0$ for all $j$;
- $\rho_\phi = -\ln\left(0.05\right) / \max\left\{||\mathbf{s}_i - \mathbf{s}_j||\right\}$;
- Gaussian specific: $\sigma^2_\epsilon = \text{variance}(\mathbf{Y})$;
- Negative binomial specific: $r = 100$.

## Additional Information

- v: Required input; an $\sum_{j=1}^m n_j$-length vector with the location number that the observation is connected to (i.e., $1, \dots, m$);
- It is recommended to scale the matrix of spatial distances by the largest observed distance in the matrix prior to running the model.
- The full_dists matrix is a $m + k$ by $m + k$ matrix where the upper left $m$ by $m$ matrix describes the distances between the $m$ observed locations, the bottom right $k$ by $k$ matrix describes the distances between the selected grid of $k << m$ locations (if the computational version is used), and the off-diagonal matrices describe the distances between the observed and grid locations (i.e., this full distances matrix is computed after stacking the observed locations and sampled grid locations into a single vector). If the computational version is not needed, input the observed locations instead of the sampled locations when calculating these distances.